91, 11943–11947 (1994).
11. Grimm, D., Kern, A., Rittner, K. & Kleinschmidt, J.A. Novel tools for production and purification of recombinant adenoassociated virus vectors. *Hum. Gene Ther.* **9**, 2745–2760 (1998).
12. Xiao, X., Li, J. & Samulski, R.J. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J. Virol.* **72**, 2224–2232 (1998).
13. Zolotukhin, S. *et al.* Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther.* **6**, 973–985 (1999).
14. Miller, A.D. *et al.* Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus. *J. Virol.* **65**, 2220–2224 (1991).
15. DuBridge, R.B. *et al.* Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol. Cell. Biol.* **7**, 379–387 (1987).
16. Inoue, N. & Russell, D.W. Packaging cells based on inducible gene amplification for the production of adeno-associated virus vectors. *J. Virol.* **72**, 7024–7031 (1998).
17. Hawley, R.G., Lieu, F.H., Fong, A.Z. & Hawley, T.S. Versatile retroviral vectors for potential use in gene therapy. *Gene Ther.* **1**, 136–138 (1994).
18. Levitt, N., Briggs, D., Gil, A. & Proudfoot, N.J. Definition of an efficient synthetic poly(A) site. *Genes Dev.* **3**, 1019–1025 (1989).
19. Miller, A.D. & Rosman, G.J. Improved retroviral vectors for gene transfer and expression. *Biotechniques* **7**, 980–982 (1989).
20. Weinrich, S.L. *et al.* Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. *Nat. Genet.* **17**, 498–502 (1997).

Amine-modified random primers to label probes for DNA microarrays

Charlie C. Xiang[1], Olga A. Kozhich[3], Mei Chen[1], Jason M. Inman[3], Quang N. Phan[1], Yidong Chen[2], and Michael J. Brownstein[1]*

DNA microarrays have been used to study the expression of thousands of genes at the same time in a variety of cells and tissues[1–3]. The methods most commonly used to label probes for microarray studies require a minimum of 20 μg of total RNA or 2 μg of poly(A) RNA[4,5]. This has made it difficult to study small and rare tissue samples. RNA amplification techniques and improved labeling methods have recently been described[6–9]. These new procedures and reagents allow the use of less input RNA, but they are relatively time-consuming and expensive. Here we introduce a technique for preparing fluorescent probes that can be used to label as little as 1 μg of total RNA. The method is based on priming cDNA synthesis with random hexamer oligonucleotides, on the 5′ ends of which are bases with free amino groups. These amine-modified primers are incorporated into the cDNA along with aminoallyl nucleotides, and fluorescent dyes are then chemically added to the free amines. The method is simple to execute, and amine-reactive dyes are considerably less expensive than dye-labeled bases or dendrimers.

Gene expression profiling relies on high-density, ordered DNA arrays and labeled probes. Progress in printing high-quality arrays has been rapid because of improvements in materials and fabrication techniques[10,11], but the development of simple, robust, and inexpensive methods for labeling probes has lagged behind. We describe a method that allows the preparation of fluorescent probes from as little as 1 mg of tissue. Previous methods for generating such probes are based on the incorporation of either dye-labeled nucleotides or aminoallyl nucleotides into reverse-transcribed DNA[12,13]. Like others, we used total RNA instead of poly(A) RNA because the former is easy to prepare in good yield using simple, standard protocols. We primed cDNA synthesis with random hexamers and incorporated aminoallyl-dUTP into the products. The resulting probes gave stronger signals than did probes generated by oligo-dT priming (data not shown). We further improved the signal strength by placing amino C6dT (thymidine modified at the 5 position with an 8–9 carbon chain ending in a primary amine) on the 5′ end of each hexamer; this allows dye molecules to be added to amino groups on bases in both the cDNAs and the incorporated primers. This approach produced substantially more fluorescence as compared with random hexamer priming (Fig. 1A).

The use of amino C6dT-modified random hexamers gave results comparable to those obtained using the conventional technique of directly incorporating dye-labeled bases into oligo-dT–primed cDNA, despite the fact that the method required one tenth as much total RNA. When optimal amounts of total RNA were used to make probes, the average signal strengths, noise levels, and signal-to-noise ratios for the two techniques were similar even though our method labels products from tRNAs and rRNAs that must be washed off the arrays (Fig. 1B).

To determine whether the two methods would reveal the same set of differentially expressed genes when used to profile two cell lines, we labeled probes by means of oligo-dT priming and direct incorporation of bases labeled with the dyes Cy5 and Cy3 using 50 μg of total RNA from mouse C2 myoblast and 3T3 fibroblast cells, respectively. Three labeled products were prepared from each total RNA sample, and pairs of probes (C2 + 3T3) were hybridized to three separate 9,568-element mouse cDNA arrays. The same procedure was carried out in parallel using 5 μg samples of C2 and 3T3 total RNA labeled with our protocol. When we searched for genes that were threefold over- or underexpressed by the two cell lines, we found 99 genes with the conventional method, and 102 with our method. Between the two groups, 95 genes were the same, and these had remarkably similar expression ratios (Table 1). When we looked for genes that were twofold over- or underexpressed, the total number of genes detected by both methods climbed to 298, but the number of nonoverlapping species also increased to 56 for the conventional method, and to 80 for our method. This may reflect noise in the system (see later) or differences in the abilities of the two methods to label certain gene products.

To determine the minimal amount of total RNA required for probe labeling with our method, we first undertook "self-on-self" experiments, comparing the expression profiles of serial dilutions of C2 RNA (labeled with Cy3) to the profile seen with 5 μg of the same RNA (labeled with Cy5). The profiles are very similar for amounts of total RNA down to 1 μg. The results for 5 μg vs. 5 μg and 5 μg vs. 1 μg are illustrated in Figure 2.

We next conducted a more stringent analysis. Because 95 genes are threefold over- or underexpressed when C2 and 3T3 cell profiles are compared using an optimal amount of total RNA, we asked how many of these genes remained threefold changed as progressively smaller amounts of RNA are labeled. We diluted the C2 and 3T3 RNA samples in parallel, labeled them with Cy3 and Cy5, respectively, mixed the products, and probed one 9,568-element array per dilution. We found most of the 95 differentially expressed genes (for which ratios between signals from the two cell lines were ≥3 or ≤1/3) when we labeled 5 μg (95 genes), 2.5 μg (90 genes), and 1 μg (87 genes) of total RNA. The number of genes that were not on the list but that were threefold changed (an average of 12) was fairly small. With 0.5 μg of RNA, only 72 of the differentially expressed genes were found, but the number of extraneous genes, 11, remained low. With 0.25 μg or 0.1 μg of RNA, there was a further decrease in differentially expressed genes

[1]*Laboratory of Genetics, National Institute of Mental Health and* [2]*Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892.* [3]*SAIC Frederick, Frederick, MD 21702. *Corresponding author (mike@codon.nih.gov).*

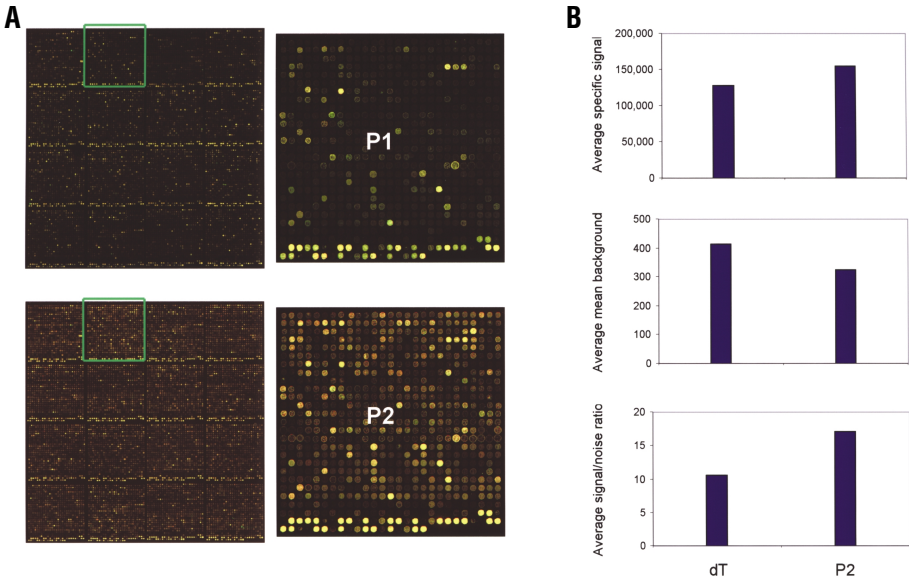

Figure 1. Results obtained with three probe labeling methods. (A) Comparison of cDNA probes reverse-transcribed from total RNA in the presence of aminoallyl-dUTP with random hexamer (P1) or amine-modified random hexamer (P2) priming. We used 5 μg of mouse C2 RNA for both Cy3 and Cy5 labelings (Experimental Protocol) and combined the products. The P1 and P2 probes were hybridized to two separate 9,568-element arrays on a single glass slide, and arrays were scanned at the same laser power and PMT settings (620 and 600 for the Cy5 and Cy3 channels, respectively). Images of the entire arrays and 720-element subarrays are shown. The P2-primed probe gave substantially higher signals than did the P1-primed probe. (B) Direct incorporation of dye-labeled bases driven by oligo-dT (dT) vs. indirect labeling with aminoallyl-dUTP incorporation driven by amine-modified random primer (P2). The direct labeling method required 50 μg of mouse C2 total RNA; our method used 5 μg of the same RNA. Data shown are from the Cy3 (green) channel, but results seen in the Cy5 (red) channel were quite similar. Two arrays on the same slide were probed with the dT and P2 products, and scanned at a PMT setting of 650. The ArraySuite DeArray subroutine was used to analyze the data. Our method gave a somewhat higher average signal, lower average background, and higher signal-to-noise ratio than the conventional one even though we used one tenth as much input RNA.

(53 and 58, respectively) and a marked increase in illegitimate genes (71 and 97, respectively). This study seemed to confirm the results of the self-on-self experiment, showing that our method can be used to label as little as 1 μg of total RNA. To reconfirm this, we compared three arrays developed with C2 and 3T3 probes made from 5 μg of RNA to three arrays developed with probes made from 1 μg of RNA. As expected, the results were quite similar (data not shown).

Our technique for preparing fluorescent probes using amine-modified random hexamer primers is relatively simple and inexpensive, and requires as little as 1 μg of RNA. It permits the preparation of hybridization probes from total RNA extracted from ~1 mg of tissue without amplification, and should be useful for routine array studies. Coupled with RNA amplification, the method should allow profiling of a few tens of cells and, in tandem with dendrimer labeling or tyramide-based detection schemes, even single cells.

Experimental protocol

Arrays. Mouse cDNA microarrays with 9,568 elements were printed on poly-L-lysine-coated glass slides using an OmniGrid arrayer (GeneMachines, San Carlos, CA). The methods used to make cDNA microarrays, including slide coating with poly-L-lysine, array fabrication, and post-processing, can be found on the websites of Stanford University[13] and the National Human Genome Research Institute (NHGRI)[14]. The cDNA clones were obtained from Bento Soares (University of Iowa, Iowa City, IA) and are part of the BMAP (Brain Molecular Anatomy Project) set[15]. The authors may be contacted for a complete list of the cDNAs used.

RNA extraction. Total RNAs from mouse C2 and NIH3T3 cells were extracted using TRIzol reagent (Invitrogen/Life Technologies, Carlsbad, CA) according to the manufacturer's instructions.

Reagents. The following reagents and materials were used to label probes: RNasin RNase inhibitor (Promega, Madison, WI); Superscript II reverse transcriptase, 5× first-strand buffer, 0.1 M dithiothreitol (DTT; Invitrogen/Life Technologies); Microcon 30 concentrator (Millipore, Bedford, MA); monofunctional NHS-ester Cy3 and Cy5 dyes and dNTPs (Amersham Pharmacia, Piscataway, NJ); QIAquick PCR purification kit and MinElute PCR purification kit from Qiagen (Valencia, CA); (5-[3-aminoallyl]-2-deoxyuridine 5-triphosphate (aa-dUTP); Sigma, St. Louis, MO); oligo-dT and random hexamer primers (Invitrogen/Life Technologies). Custom-synthesized amine-modified random primers were purchased from Sigma Genosys (The Woodlands, TX).

Probe labeling. For direct labeling of probes with oligo-dT primer, the protocol of the Brown Lab at Stanford[13] was adopted; 50 μg of total RNA was used as starting material. For labeling with our method we proceeded as follows: 0.1–5 μg of total RNA (15.5 μl) was combined with amine-modified random primer (2 μg/μl, 2 μl) and RNase inhibitor (5 units/μl, 1 μl). The mix was incubated at 70°C for 10 min, and then chilled on ice for 10 min. Primer–RNA solution was added to the reverse transcriptase mix (5× first-strand buffer, 6 μl; 50× aa-dUTP/dNTPs (25 mM dATP, dGTP, and dCTP, 15 mM dTTP, and 10 mM aminoallyl-dUTP), 0.6 μl; DTT, 0.1 M, 3 μl; Superscript II reverse transcriptase (Invitrogen/Life Technologies), 2 μl) and incubated at 42°C for 2 h. The reaction was terminated by adding EDTA (0.5 M, 10 μl), and the RNA was hydrolyzed with NaOH (1 M, 10 μl) at 65°C for 30 min.

The solution was neutralized with HCl (1 M, 10 μl), and then MinElute

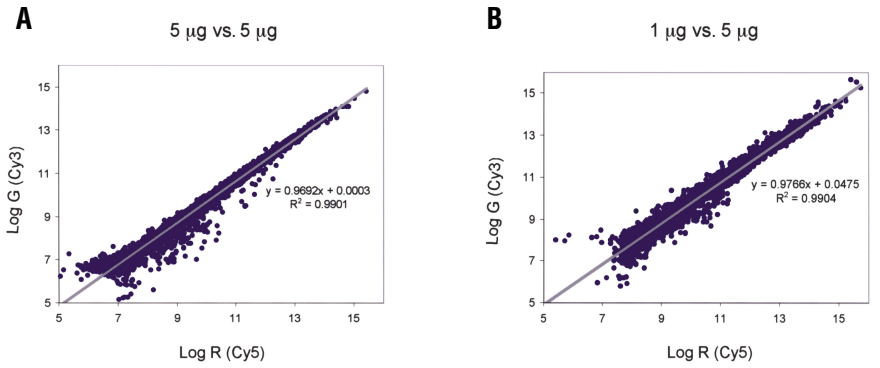

Figure 2. Labeling of the same RNA sample with Cy5 and Cy3. (A) We used our method to prepare Cy5 and Cy3 probes with 5 μg of C2 total RNA per labeling. We combined the two probes, hybridized them to arrays, and scanned them (PMT voltages of 600 and 550 for Cy5 and Cy3, respectively). Scatter plots of log Cy5 vs. log Cy3 are shown. There was a strong correlation between the signals in the two channels. (B) Cy5 and Cy3 probes were prepared from 5 μg and 1 μg of total C2 RNA, respectively. PMT voltages of 600 and 580 were used to scan the Cy5 and Cy3 channels. The red and green signals were strongly correlated. All data points are shown.

Table 1. A list of 95 genes that are differentially expressed in mouse 3T3 vs. C2 cells (3T3/C2 ratios ≥3 or ≤1/3)[a]

GeneBank	Unigene	dTa	dTb	dTc	P2a	P2b	P2c	Clone description
AI849214	Mm.105330	18.3218	17.5265	16.2183	20.7580	25.5790	22.7692	Whey acidic protein
AI848293	Mm.34507	6.9711	5.3423	3.5613	8.1450	6.4472	6.7642	ESTs
AI847098	Mm.29982	5.2438	4.5541	4.5709	5.6932	5.1469	5.7707	ERO1-like (*S. cerevisiae*)
AI852317	Mm.4063	3.7839	3.5895	4.4260	5.4368	5.0830	5.2313	N-myc downstream regulated 1
AI844828	Mm.2834	3.7150	3.7105	3.9967	5.1164	4.9883	4.7269	Glycine transporter 1
AI846827	Mm.70667	5.2250	4.0641	3.4577	4.6474	4.2576	4.2443	*Mus musculus*, similar to oxidation resistance 1
AI843085	Mm.157648	5.5280	4.4538	4.7526	4.5177	4.0203	4.3967	RIKEN cDNA 5730403B10 gene
AI842716	Mm.140158	5.5015	5.8588	5.1314	4.4732	4.4336	4.4295	Cytochrome P450, 51
AI836864	Mm.4704	6.6261	5.7066	3.7317	4.4534	3.9560	4.4178	Forkhead box G1
AI853347	Mm.21884	4.0523	3.9847	3.3449	4.4364	4.7968	4.3672	ESTs, weakly similar to GTPase-activating protein SPA-1
AI843677	Mm.45357	3.7376	3.5943	3.5423	3.8309	3.4541	3.3920	Erbb2 interacting protein
AI838612	Mm.14601	3.0926	3.3974	3.2623	3.6027	3.4499	3.4159	Glutathione *S*-transferase, mu 2
AI848205	Mm.35844	3.6669	3.3875	3.1423	3.4911	3.0100	3.1019	Growth arrest specific 5
AI850589	Mm.22627	3.7784	3.1037	3.1616	3.2339	3.2407	3.6818	Epidermal growth factor receptor pathway substrate 15
AI852765	Mm.24193	0.3300	0.3343	0.3183	0.3254	0.2847	0.3249	Glypican 1
AI836264	Mm.4871	0.1492	0.1253	0.1183	0.3200	0.3100	0.2357	Tissue inhibitor of metalloproteinase 3
AI844851	Mm.10406	0.3209	0.3235	0.2910	0.3243	0.2993	0.3025	RIKEN cDNA 3110001M13 gene
AI851985	Mm.29586	0.2668	0.2559	0.2278	0.3233	0.2814	0.3107	RIKEN cDNA 2610024P12 gene
AI845475	Mm.30811	0.1031	0.1333	0.1210	0.3180	0.3200	0.3100	ESTs
AI853172	Mm.27173	0.2968	0.3133	0.3032	0.3132	0.2847	0.3100	Ectoplacental cone sequence
AI835858	Mm.27685	0.2834	0.2925	0.2512	0.3114	0.3067	0.2751	ESTs, Highly similar to tropomyosin 4 (*Rattus norvegicus*)
AI836045	Mm.29976	0.2461	0.3202	0.2812	0.3016	0.3236	0.2702	Septin 5
AI843823	Mm.7414	0.1481	0.1690	0.1445	0.2971	0.3129	0.2507	Neuron-specific gene family member 1
AI844342	Mm.182255	0.1773	0.2039	0.2446	0.2833	0.3164	0.3083	CD97 antigen
AI835331	Mm.544	0.2802	0.3336	0.3057	0.2829	0.1995	0.2646	Phosphoprotein enriched in astrocytes 15
AI845602	Mm.4146	0.2438	0.2668	0.3188	0.2727	0.2349	0.2469	Platelet-derived growth factor receptor, β-polypeptide
AI838302	Mm.4426	0.2816	0.2966	0.3223	0.2702	0.2466	0.2872	Cd63 antigen
AI835546	Mm.3117	0.2023	0.2238	0.2903	0.2696	0.3022	0.3240	T-cell death-associated gene
AI853531	Mm.21679	0.2340	0.3006	0.3272	0.2691	0.2573	0.2708	RIKEN cDNA 1300002F13 gene
AI842302	Mm.4139	0.3176	0.3029	0.3261	0.2652	0.2259	0.2783	Rhotekin
AI835620	No data	0.2793	0.3169	0.3180	0.2637	0.2298	0.2679	No data
AI845774	Mm.856	0.2799	0.2757	0.3172	0.2630	0.2362	0.2575	Transmembrane 4 superfamily member 1
AI838659	Mm.262	0.2496	0.2866	0.3001	0.2484	0.2192	0.2592	*ras* homolog gene family, member C
AI848618	Mm.29010	0.1939	0.2150	0.2075	0.2473	0.2205	0.2216	Membrane-bound C2 domain–containing protein
AI851997	Mm.29010	0.2759	0.2851	0.3298	0.2462	0.2379	0.2648	Membrane-bound C2 domain–containing protein
AI852812	Mm.2308	0.2209	0.2669	0.3063	0.2409	0.2236	0.2485	Hemoglobin Z, β-like embryonic chain
AI844356	Mm.1017	0.2547	0.2658	0.2582	0.2261	0.2191	0.2255	Esterase 10
AI851647	Mm.22240	0.2365	0.2571	0.2440	0.2219	0.2185	0.2236	ESTs, weakly similar to SH3BGR protein
AI838551	Mm.2792	0.1605	0.1832	0.1807	0.2191	0.1398	0.2238	Prostaglandin-endoperoxide synthase 1
AI842654	Mm.8180	0.2336	0.2595	0.2941	0.2182	0.2249	0.2627	Lymphocyte antigen 6 complex
AI841122	Mm.39804	0.2427	0.2581	0.3048	0.2139	0.2408	0.2015	EST
AI838653	Mm.181074	0.2615	0.2885	0.3198	0.2073	0.2179	0.2407	RIKEN cDNA 2610001E17 gene
AI838959	Mm.16537	0.1483	0.1504	0.2370	0.2014	0.2943	0.2463	Actin, α-2, smooth muscle, aorta
AI842847	Mm.8245	0.2013	0.2803	0.2512	0.1975	0.1770	0.1926	Tissue inhibitor of metalloproteinase
AI838351	No data	0.1422	0.1998	0.0999	0.1913	0.3317	0.2076	No data
AI837390	Mm.43278	0.1418	0.1444	0.1499	0.1882	0.2873	0.2535	Olfactomedin-related ER-localized protein
AI844326	Mm.194675	0.2317	0.2675	0.2290	0.1847	0.0958	0.1462	EST
AI839057	No data	0.2107	0.2988	0.2685	0.1806	0.2179	0.2184	No data
AI838085	Mm.687	0.1668	0.1773	0.2450	0.1781	0.2298	0.2301	*Aplysia ras*-related homolog B (RhoB)
AI837494	Mm.39836	0.1604	0.1709	0.2824	0.1768	0.1658	0.1247	ESTs, similar to T14318 ubiquitin-protein ligase E3-α
AI836532	Mm.196484	0.1481	0.1464	0.1405	0.1645	0.1642	0.1756	EST AA408841
AI835609	Mm.1956	0.0364	0.0776	0.0791	0.1608	0.2416	0.1599	Neurofilament, light polypeptide
AI842984	Mm.980	0.1258	0.1350	0.1376	0.1602	0.2456	0.1732	Tenascin C
AI849378	Mm.2769	0.1639	0.1670	0.1944	0.1545	0.1712	0.2004	MARCKS-like protein
AI839275	Mm.738	0.1356	0.1868	0.2704	0.1503	0.2651	0.1883	Procollagen, type IV, α-1
AI844626	Mm.29975	0.0684	0.1024	0.1284	0.1489	0.1956	0.1716	RIKEN cDNA 1810003P21 gene
AI835201	Mm.8739	0.1115	0.1536	0.1402	0.1454	0.1709	0.1867	Sarcoglycan, epsilon
AI844312	Mm.3091	0.1443	0.2400	0.2183	0.1432	0.2094	0.1778	Epsin 1
AI841755	Mm.687	0.1340	0.1510	0.1345	0.1427	0.1610	0.1485	*Aplysia ras*-related homolog B (RhoB)
AI838813	Mm.192516	0.1338	0.1664	0.1652	0.1416	0.1249	0.1655	EST
AI839735	Mm.37751	0.1409	0.1558	0.1463	0.1403	0.1138	0.1486	ESTs
AI837031	Mm.157662	0.0520	0.0994	0.1407	0.1260	0.0776	0.0931	Synaptotagmin 13
AI840673	Mm.29924	0.0846	0.0945	0.1128	0.1237	0.1111	0.1437	ADP-ribosylation-like factor 6 interacting protein
AI841538	Mm.41009	0.1166	0.1329	0.2839	0.1210	0.1168	0.1004	Nedd4 WW-binding protein 4
AI847958	Mm.20246	0.1447	0.1526	0.2049	0.1173	0.0934	0.1017	RIKEN cDNA 2410004D18 gene
AI840633	Mm.38021	0.0477	0.1194	0.1215	0.1122	0.0823	0.0391	Carbohydrate (keratan sulfate Gal-6) sulfotransferase 1
AI843323	Mm.3900	0.1334	0.1957	0.2642	0.1120	0.0902	0.1358	Latent transforming growth factor–β binding protein 2
AI849869	Mm.34113	0.1241	0.1336	0.1955	0.1120	0.1015	0.1198	VPS10 domain receptor protein SORCS 2
AI840335	Mm.39154	0.0928	0.1347	0.1007	0.1104	0.1833	0.1133	EST
AI840972	Mm.29580	0.2618	0.3083	0.3024	0.1059	0.1794	0.1681	Superior cervical ganglia, neural specific 10

AI847162	Mm.29357	0.0973	0.0696	0.2018	0.1050	0.1264	0.1312	RIKEN cDNA 1300017C10 gene
AI843174	Mm.29924	0.1284	0.1426	0.1479	0.1044	0.1134	0.1473	ADP-ribosylation-like factor 6 interacting protein
AI839366	Mm.28947	0.0651	0.1159	0.1742	0.1021	0.1278	0.1395	ESTs
AI840692	No data	0.1394	0.1457	0.1741	0.0917	0.1456	0.1644	No data
AI835703	Mm.29975	0.0961	0.0827	0.0714	0.0868	0.1302	0.1381	RIKEN cDNA 1810003P21 gene
AI836865	Mm.44102	0.0503	0.0643	0.1129	0.0842	0.1727	0.1572	ESTs
AI842983	Mm.192586	0.0702	0.1325	0.1346	0.0785	0.1555	0.1091	EST
AI839950	Mm.3126	0.0492	0.0610	0.0989	0.0781	0.2076	0.1304	Four-and-a-half LIM-domain 1
AI844604	Mm.3126	0.1263	0.1328	0.1465	0.0750	0.0188	0.0613	Four-and-a-half LIM-domain 1
AI836826	Mm.2976	0.0747	0.0764	0.0755	0.0747	0.0918	0.0759	Glycoprotein 38
AI850497	Mm.41072	0.1133	0.1862	0.2509	0.0743	0.1009	0.0891	ESTs, similar to LOX5 mouse arachidonate 5-lipoxygenase
AI835403	Mm.142729	0.0965	0.1012	0.1014	0.0620	0.0778	0.0579	Thymosin, β-4, X chromosome
AI848096	Mm.17951	0.1483	0.1711	0.1888	0.0580	0.1324	0.1233	Erythrocyte protein band 4.1-like 3
AI843282	Mm.181021	0.0955	0.1120	0.1453	0.0529	0.0995	0.1095	Procollagen, type IV, α-2
AI842554	Mm.192583	0.0577	0.0889	0.0962	0.0428	0.1088	0.0815	ESTs
AI842681	Mm.20904	0.0702	0.0488	0.1056	0.0405	0.0375	0.0487	Cartilage-associated protein
AI835976	Mm.17951	0.0491	0.0372	0.0362	0.0392	0.0591	0.0431	Erythrocyte protein band 4.1-like 3
AI836468	Mm.30059	0.0495	0.0491	0.1289	0.0345	0.0690	0.0530	Myristoylated alanine-rich protein kinase C substrate
AI844038	Mm.7919	0.0322	0.0323	0.0399	0.0339	0.0511	0.0232	HGF-regulated tyrosine kinase substrate
AI838614	Mm.14802	0.0412	0.0399	0.0281	0.0331	0.0407	0.0464	H19 fetal liver mRNA
AI849859	Mm.3126	0.0204	0.0173	0.0375	0.0323	0.0641	0.0296	Four-and-a-half LIM-domain 1
AI837752	Mm.43278	0.0346	0.0221	0.0848	0.0314	0.0460	0.0454	Olfactomedin-related ER-localized protein
AI841798	Mm.4871	0.0533	0.0983	0.1831	0.0273	0.0217	0.0219	Tissue inhibitor of metalloproteinase 3
AI838607	Mm.4159	0.0277	0.0301	0.0276	0.0268	0.0559	0.0602	Thrombospondin 1
AI842703	Mm.147387	0.0284	0.0297	0.0391	0.0200	0.0205	0.0253	Procollagen, type III, α-1

[a]The table shows the results of six array experiments. Three 9,568-element arrays were developed with oligo-dT primed probes, and three others were developed with amine-modified hexamer primed probes. We used 50 μg of RNA for each oligo-dT primed labeling and 5 μg for each modified hexamer primed labeling. (See text and Experimental Protocol for details.) Array images were analyzed using ArraySuite software. Low-quality ratios were filtered as described in the Experimental Protocol. Note that the ratios of the 95 genes that are differentially expressed are quite consistent among all six experiments. Elements representing the RhoB and four-and-a-half LIM-domain 1 transcripts were printed two and three times on the array, respectively. These genes appear to be expressed at a higher level in C2 cells than in 3T3 cells, and it is reassuring that all elements representing them showed similar ratios. The four-and-a-half LIM-domain protein is known to be made in cardiac and skeletal muscle.

PCR purification kits were used to purify the cDNA: microcentrifuge tubes were filled with 300 μl buffer PB, to which 60 μl of the neutralized reaction solution was then added. MinElute columns (Qiagen) were placed in 2 ml collection tubes and samples were applied to columns, after which tubes were centrifuged for 1 min. For maximum recovery, all traces of the samples were transferred to the column. The flowthrough was poured back into the columns, which were centrifuged again for 1 min. Flowthrough was then discarded and columns were placed back into original collection tubes. After adding 750 μl of buffer PE to each of the MinElute columns, the columns were incubated for 5 min at room temperature and centrifuged for 1 min. The flowthrough was discarded and the columns were put back in the same tubes and centrifuged for an additional 1 min at 16,000*g*. MinElute columns were transferred to clean 1.5 ml microcentrifuge tubes. To elute the DNA, 10 μl of H_2O (pH 7.0–8.5) was pipetted onto the center of the membrane; columns were left at room temperature for 1 min and were then centrifuged for 5 min. Average eluate was 9 μl out of the 10 μl applied. The DNA was eluted twice more with 10 μl of H_2O; a total of 27 μl of purified cDNA solution was collected.

Sodium bicarbonate (3 μl of 1 M, pH 9.3) was added to the cDNA solution, followed by 1 μl of dye (NHS-ester Cy3 or Cy5, 62.5 μg/μl in dimethyl sulfoxide). The resulting solution was mixed by pipetting up and down several times; tubes were wrapped in aluminum foil and incubated at room temperature for 1 h in an orbital shaker (USA Scientific, Ocala, FL). The labeling reaction was stopped with 4.5 μl of 4 M hydroxylamine hydrochloride. Afterward, the tubes were vortexed, briefly centrifuged, and incubated for 30 min at room temperature in the dark.

Probe purification. Probes were cleaned with a Qia-quick PCR purification kit (Qiagen); the Cy3- and Cy5-labeled products were combined and 30 μl of water was added, followed by 500 μl of buffer PB. The samples were applied to Qia-quick columns, which were centrifuged at 13,000 rpm (16,000*g*) for 1 min, after which the flowthroughs were discarded. To wash the columns, 750 μl of buffer PE was added, columns were spun again for 1 min, and the flowthroughs were thrown away. The wash step was repeated once more, and columns were spun again to remove residual ethanol. Fresh collection tubes were placed beneath each column, 30 μl of buffer EB was added, and tubes were incubated for 1 min at room temperature. Columns were then centrifuged at 13,000 rpm (16,000*g*) for 1 min, and the elution step was repeated once. Eluates were partially dried in a vacuum centrifuge and the volumes were adjusted to 23 μl with water.

Hybridization and wash conditions. 4.5 μl of 20× saline–sodium citrate (SSC) was added along with 2 μl of poly(A) (10 mg/ml), and 0.6 μl of 10% (wt/vol) SDS, and the probes were denatured at 100°C for 2 min. The products were pipetted onto arrays, coverslips were applied, and the slides were placed in a hybridization chamber (Corning, Corning, NY). Arrays were incubated in a 65°C water bath for 16–24 h, and subsequently washed with 0.5× SSC, 0.01% (wt/vol) SDS, followed by 0.06× SSC, at room temperature for 10 min each. Slides were next placed in 50 ml tubes and spun for 5 min at 800 rpm (130*g*) at room temperature.

Array scanning. Arrays were read with a GenePix 4000A scanner (Axon, Foster City, CA) at 10 μm resolution and variable photomultiplier tube (PMT) voltage settings to obtain the maximal signal intensities with <1% (wt/vol) probe saturation. The resulting images were analyzed using IPLab (Fairfax, VA) and ArraySuite (NHGRI, Bethesda, MD) software.

Analysis. To determine the reliability of each ratio measurement, a set of quality indicators was used. To be considered reliable, intensity measurements had to satisfy the following criteria: (i) association of a sufficiently large number of pixels with the element, (ii) flat local background, (iii) uniform signal consistency within the target area, and (iv) unsaturation of the majority of the signal pixels. For each ratio measurement R/G, one further condition was imposed—an average signal (R + G)/2 that is at least three times the noise level. A detailed discussion of this method is given by Chen *et al.*[16,17].

To analyze the consistency of over- or underexpressed genes, we asked the following question: given the result that 95 genes were consistently expressed in all of the three replica experiments, how many genes will survive the comparison when a fourth microarray is examined using the same experimental conditions? To answer this question, we studied a model in which a log-transformed gene-expression ratio, $w = \log t$, is assumed to be normally distributed with a standard deviation of σ. For this model, the probability of observing a ratio measurement >3.0 in one experiment is

$$p = P_{\mu=w}(x > \ln 3) = \int_{\ln 3}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-w)^2}{2\sigma^2}} dx \qquad (1)$$

where ln denotes the natural logarithm. The probabilities that a ratio measurement would be >3 in all of two, three, or four experiments are simply p^2, p^3, and p^4, respectively. We further assumed that within a confined ratio

region $[l_1, l_2]$, where $l_1 \le 3 \le l_2$, there is equal probability for all ratio values, or p_r. Thus, the probability that any gene ratio within the region l_1 to l_2 is >3 is given by

$$p = \int_{l_1}^{l_2} p_r P_{\mu=w}(x > \ln 3) dw = p_r \int_{w=l_1}^{l_2} \int_{x=\ln 3}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-w)^2}{2\sigma^2}} dx\, dw \qquad (2)$$

The difference in the expected number of genes in three consistent experiments and four consistent experiments is

$$n = N\left[\int_{l_1}^{l_2} p_r \left[P_{\mu=w}(x > \ln 3)\right]^3 dw - \int_{l_1}^{l_2} p_r \left[P_{\mu=w}(x > \ln 3)\right]^4 dw\right] \qquad (3)$$

where N is the total number of genes within the region of $[l_1, l_2]$. The result for expression ratios <1/3 can be similarly derived. Given that the number of consistent genes were known (m = 95 in this study), we have

$$n = m\left[1 - \frac{\int_{l_1}^{l_2} \left[P_{\mu=w}(x > \ln 3)\right]^4 dw}{\int_{l_1}^{l_2} \left[P_{\mu=w}(x > \ln 3)\right]^3 dw}\right] \qquad (4)$$

To numerically evaluate the above equation, we first chose a typical $\sigma = 0.07$, which was estimated from the duplicated elements printed on the array. We also selected a typical region $[l_1, l_2]$ for consideration, $[\ln 2.0, \ln 4.5]$. For m = 95 (we combined the threefold changes together, as the equations (4) for overexpression and underexpression are identical), we found n = 3.6. If instead we used $\sigma = 0.14$, which was the typical variation derived from the self-on-self experiment, we found n = 8.6. Therefore, when a fourth microarray is studied under identical experimental conditions, among 95 consistently threefold–differentially expressed genes, we expect 4–9 genes to be dropped as a result of random variation of the microarray assay. In other words, the 90 and 87 genes seen with 2.5 μg and 1 μg of input RNA were within the expectations of this analysis. Thus they represent comparable experiments even though the amount of RNA used to make probe was different. For less input RNA (from which ≤72 genes in the differentially expressed class were detected), the number is far below that expected, and we conclude that insufficient RNA was employed. Nonetheless, the gain in stain was mainly not in vain.

Acknowledgments

This project was partly supported by federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

Competing interests statement

The authors declare competing financial interests: see the Nature Biotechnology *website (http://biotech.nature.com) for details.*

Received 18 January 2002; accepted 3 April 2002

1. DeRisi, J. *et al.* Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**, 457–460 (1996).
2. Heller, R.A. *et al.* Discovery and analysis of inflammatory disease–related genes using cDNA microarray. *Proc. Natl. Acad. Sci. USA* **94**, 2150–2155 (1997).
3. Wen, X. *et al.* Large-scale temporal gene expression mapping of central nervous system development. *Proc. Natl. Acad. Sci. USA* **95**, 334–339 (1998).
4. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
5. Duggan, D., Bitter, M., Chen, Y., Meltzer, P. & Trent, J. Expression profiling using cDNA microarrays. *Nat. Genet. Suppl.* **21**, 10–14 (1999).
6. Wang, E., Miller, L.D., Ohnmacht, G.A., Liu, E.T. & Marincola, F.M. High-fidelity mRNA amplification for gene profiling. *Nat. Biotechnol.* **18**, 457–459 (2000).
7. Stears, R.L., Getts, R.C. & Gullans, S.R. A novel, sensitive detection system for high-density microarrays using dendrimer technology. *Physiol. Genomics* **3**, 93–99 (2000).
8. Baelde, H.J. *et al.* High quality RNA isolation from tumors with low cellularity and high extracellular matrix component for cDNA microarrays: application to chondrosarcoma. *J. Clin. Pathol.* **54**, 778–782 (2001).
9. Wong, K.-K., Cheng, R.S. & Mok, S.C. Identification of differentially expressed genes from ovarian cancer cells by MICROMAX cDNA microarray system. *Biotechniques* **30**, 670–675 (2001).
10. Eisen, M.B. & Brown, P.O. DNA arrays for analysis of gene expression. *Methods Enzymol.* **303**, 179–205 (1999).
11. Lashkari, D.A. *et al.* Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. USA* **94**, 13057–13062 (1997).
12. Nimmakayalu, M., Henegariu, O., Ward, D.C. & Bray-Ward, P. Simple method for preparation of fluor/hapten-labeled dUTP. *Biotechniques* **28**, 518–522 (2000).
13. Stanford University School of Medicine, Brown lab. *The MGuide*. http://cmgm.stanford.edu/pbrown/mguide/index.html.
14. National Human Genome Research Institute Division of Intramural Research. *Microarray Project (uAP)*. http://www.nhgri.nih.gov/DIR/Microarray/main.html.
15. The University of Iowa. *BMAP Project*. http://brainest.eng.uiowa.edu.
16. Chen, Y., Dougherty, E.R. & Bittner, M.L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Biomedical Optics* **2**, 364–374 (1997).
17. Chen, Y. *et al.* Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics*, in press (2002).